ANSWERS AND EXPLANATIONS

Practice Examination 3

Section I: Multiple Choice Solutions

1. D

   This is a χ^2 goodness-of-fit test. The null hypothesis is that the distribution of seniors in the community fits the distribution of senior citizens in the United States. The alternative hypothesis is that the distribution of seniors in the community does not fit. The expected counts are as follows.

Age	65–74	75–84	85–94	95 and over
Percentage	363.08	201.57	56.34	5.008

$$\begin{aligned}\chi^2 &= \frac{(300 - 363.08)^2}{363.08} + \frac{(232 - 201.57)^2}{201.57} \\ &\quad + \frac{(92 - 56.34)^2}{56.34} + \frac{(2 - 5.008)^2}{5.008} \\ &\approx 39.930\end{aligned}$$

   With 3 degrees of freedom, the resulting p-value is $1.1026 \times 10^{-8} \approx 0 < 0.001$. Therefore, the results are statistically significant at the 0.001 level.

2. C

   A scatterplot is an appropriate graphical display to compare two quantitative variables. For this problem, there is one quantitative variable split into two categories.

3. D

   The population size is 500, so a sample size of either 50 or 30 is appropriate. We are not given σ, so a t-test is appropriate. $df = n - 1$, where n = sample size. If $n = 30$, $df = 29$.

4. E

   The mean of $X + Y = E(X + Y) = E(X) + E(Y) = 2.5 + 4.7 = 7.2$. Because the random variables are independent, we can add the *variances*, but *not* the standard deviations, so

   $\text{Var}(X) = (0.3)^2$, $\text{Var}(Y) = (0.4)^2$, and

   $\text{Var}(X + Y) = 0.09 + 0.16 = 0.25$.

   The standard deviation is $\sqrt{0.25} = 0.5$.

5. E

Random variables may take on any values; for example, a loss of \$2 at the racetrack may be represented by the random variable value (-2). Expected value to a player at a casino is negative (otherwise the "house" would be losing money). Since the variance is computed by the formula $\text{Var}(X) = E(X - \mu)^2P(x)$, the value will be positive or zero (nonnegative). The variance is zero if the difference between the random variable and the mean is zero.

6. D

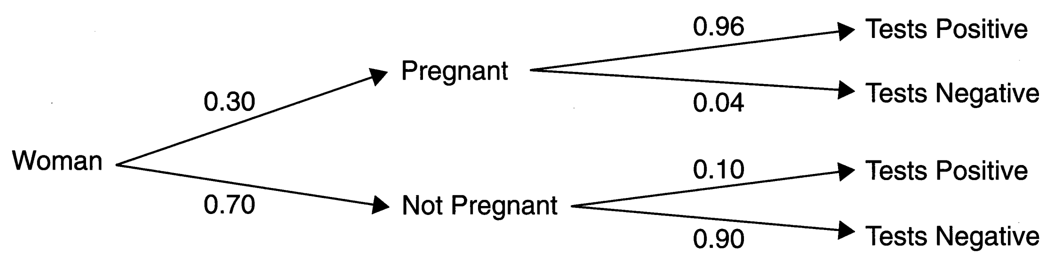
The formula for the construction of the confidence interval is

$$\hat{p} \pm z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \text{ where } \hat{p} = \tfrac{20}{100}, z^* = 1.96, \text{ and } n = 100.$$

If the hypothesized value for the proportion is contained in the 95% confidence interval, the results of the hypothesis test will not be significant. Since 0.151 is contained in the 95% confidence interval, (0.1216, 0.2784), we will not be able to show a significant difference between our sample proportion and the 2000 proportion of women in the Army.

7. C

Make a tree diagram to show the probabilities as stated in the problem.



We want to find

$$P(\text{pregnant} \mid \text{tests positive}) = \frac{P(\text{pregnant} \cap \text{tests positive})}{P(\text{tests positive})}$$

$$= \frac{(0.3)(0.96)}{(0.3)(0.96) + (0.7)(0.1)} \approx 0.8045.$$

8. C

Bar graphs are a graphical display for categorical values. Because the variable is categorical, there is no scale for the horizontal axis, and there is a gap between bars. Additionally, it doesn't make sense to assign numerical summaries to categorical data.

9. A

Although the histograms show opposite skew directions, you must be careful to read the labels on any graph. It is clear that some students in class B earn very little or no money and that none of them earns more than \$350. In class A, everyone has a job and earns at least \$50 with some earning in the \$350–\$450 range. While the centers for both distributions lie in the same modal class, these extreme values will pull their means in opposite directions (toward their respective tails). Thus, the mean for class A is higher than for class B.

10. C

The probability that Tom's car will break down is 0.5: $P(T) = 0.5$. The probability that Janice's car will break down is 0.5: $P(J) = 0.5$. The probability that both of their cars will break down is 0.3: $P(T \cap J) = 0.3$. We wish to find the probability that Tom or Janice's car will break down.

$$P(T \cup J) = P(T) + P(J) - P(T \cap J) = 0.5 + 0.5 - 0.3 = 0.7$$

11. A

The median is the value at the 50th percentile, which is 16 years of age for this problem. The interquartile range is found by taking the 75th percentile–25th percentile, in this case, $17 - 16 = 1$. Notice that the median and Q1 are the same in this problem.

12. C

$$n = \left(\frac{z^{*}\sigma}{\text{ME}}\right)^2 = \left(\frac{1.645 \cdot 10}{0.5}\right)^2$$
$$= 1082.41 \approx 1083$$

Thus, 1083 death records should be examined.

13. C

The width of a confidence interval is determined by doubling the margin of error. For the construction of a confidence interval for means, the margin of error is $t^{*}\frac{s}{\sqrt{n}}$. If the sample size is quadrupled, the new margin of error becomes

$$t^{*}\frac{s}{\sqrt{4n}} = t^{*}\frac{s}{2\sqrt{n}} = \frac{1}{2}t^{*}\frac{s}{\sqrt{n}}.$$

Therefore, if the margin of error is halved, so too is the width of the confidence interval.

14. C

A p-value outside of the context of a problem setting tells us only the probability that a test statistic at least as extreme as ours occurs when the null hypothesis is true. We cannot reject the null hypothesis without an alpha level or problem setting that suggests an appropriate level.

15. D
A simple random sample is one in which each individual has the same probability of being chosen, and each sample of size 100 has the same probability of being chosen. Only answer choice D satisfies both conditions.

16. B
The mean, range, standard deviation, and variance are not resistant to outliers. The median and interquartile range are resistant to outliers.

17. C
The population of California is much larger than that of Montana. Therefore, 43% of California's residents constitute many more individuals than 43% of Montana's residents. As a result, weighting must be considered to state an overall percentage, and the percentages cannot be added.

18. C
Plot 3 has a strong negative correlation; plot 2 has a moderate negative correlation; plot 4 has a weak positive correlation; and plot 1 has a strong positive correlation.

19. D
From the display, it is evident that 0 is the minimum value and 7 is the maximum value. There are 33 data values, so the median is in the 17th position. The median is 6. The lower quartile is between the 8th and 9th positions and has a value of 4. The upper quartile is 7. Thus, the IQR is $7 - 4 = 3$. Using the 1.5 IQR rule, 1.5 IQR $= 4.5$.

$Q1 - 1.5\,IQR = 4 - 4.5 = -0.5$

Therefore, there are no outliers on the low end.

20. D
A census is not an experiment. A parameter is a value used to describe a population. A population is the entire group of individuals we want information about, and voluntary response samples are generally biased.

21. D
We are trying to determine if the age breakdown in the city fits the distribution of ages in the country.

22. C
Only patients in a high-risk group were part of the study. The results are not readily generalizable to all people who may take this over-the-counter medication.

23. D
Power $= 1 - \beta$, where $\beta =$ the probability of a Type II error.
$82\% = 1 - \beta$, so $\beta = 18\%$.

24. D
The degrees of freedom in a one-sample situation are $df = n - 1$.

25. D

The formula for the confidence interval is

$$(\bar{x}_B - \bar{x}_A) \pm t^* \sqrt{\frac{s_B^2}{n_B} + \frac{s_A^2}{n_A}} = (90 - 88) \pm t^* \sqrt{\frac{6^2}{32} + \frac{8^2}{32}}.$$

Using technology will yield a t^*-value with 57.493 degrees of freedom. When using the t-tables, the more conservative value of 30 degrees of freedom should be used. ($32 - 1 = 31$ *df*, but 31 does not appear on the table.) The resulting interval is $(-1.54, 5.54)$. Thus, the mean for team A is between 5.54 points less than and 1.54 points more than the mean for team B.

26. D

We want to find the probability of rolling three *or* four *or* five 1s. Use the binomial probability formula

$$C(n, r)p^r(1 - p)^{n-r}$$

$$= C(5, 3)\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 + C(5, 4)\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^1 + C(5, 5)\left(\frac{1}{6}\right)^5\left(\frac{5}{6}\right)^0$$

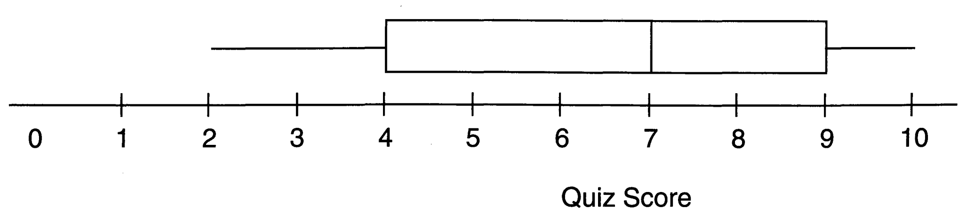
$$\approx 0.03215 + 0.00321 + 0.00012 \approx 0.0354.$$

Alternatively, you may use the TI-83/84 command

binomcdf $\left(5, \frac{1}{6}, 2\right) \approx .9645$, which finds the probability of rolling zero, one, or two 1s and subtract from 1.

27. B

The five-number summary is 2, 4, 7, 9, 10. The boxplot of the data is



28. E

A confidence interval may or may not contain the true proportion. 95% of intervals constructed with sample data like ours will contain the true proportion. A confidence interval gives a range of plausible values for the population proportion but does not tell us anything about a particular college.

29. B
There is nothing tricky here. Just follow the part of the tree that shows event A, then follow to B. When event A occurs, event B will occur 0.2 or 20% of the time. You will get the same result by using
$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{(0.25)(0.2)}{(0.25)} = 0.2,$$
but with unnecessary effort.

30. E
Studies that involve human subjects are especially sensitive to ethical issues. In this case, the drug being studied should be available to all involved in the study since the study concerned the effectiveness and the results were very positive.

31. B
The data show a positive, moderately strong linear association.

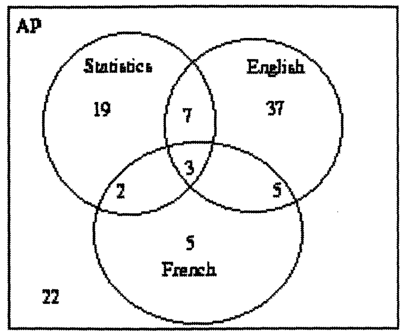
32. E
We are finding a 98% confidence interval for a mean given a sample of size 15. A t-interval with 14 degrees of freedom would be appropriate.
$$\bar{x} \pm t^*_{df}\frac{s}{\sqrt{n}} = 1.25 \pm 2.62\frac{0.39}{\sqrt{15}}$$

33. B
Since the president is interested only in assessing the parking needs of borough residents, the population of interest is borough residents who own cars.

34. E
In order to find a sample size, we could use the conservative value of 0.5 for p, but we are still missing the confidence level. Without a confidence level, we cannot find a value for z^*. To find a sample size given the margin of error in a problem dealing with proportions:
$$\text{ME} = z^*\sqrt{\frac{p(1-p)}{n}} = \frac{z^*\sqrt{p(1-p)}}{\sqrt{n}} \Rightarrow$$
$$\sqrt{n} = \frac{z^*\sqrt{p(1-p)}}{\text{ME}} \Rightarrow n = \left(\frac{z^*\sqrt{p(1-p)}}{\text{ME}}\right)^2.$$

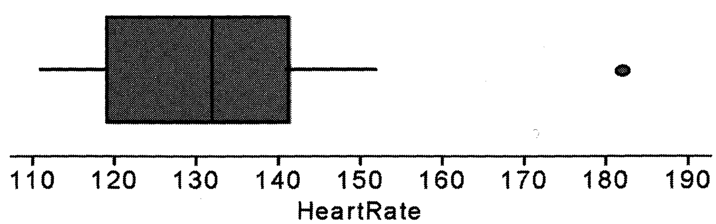
35. B
This was a comparative experiment because a treatment (aspirin/placebo) was imposed and the responses were compared. Because neither the patients nor the doctors knew who was getting which pill, the experiment was double-blind.

36. C
Make a Venn diagram to illustrate the situation. Since 22 of 100 AP* students take none of the 3 courses, the answer to the question asked is 22%.



37. B
Only statement II is correct. The two-sample t-test only allows us to make conclusions about the average student grade for each professor, not individual students. Additionally, we were told that we have random samples of grades for each professor, and the distribution of grades is approximately normal for each as well. It is reasonable to conclude that the grades of the two professors are independent. The correct interpretation of a p-value says that if the null hypothesis is true (there is no difference in grades between the two professors), then we can expect to get results as extreme as those from our sample 2.2% of the time (p-value $= 0.022$).

38. E
A modified boxplot reveals an outlier; therefore, the conditions necessary to find a 95% confidence interval for a mean with a sample of size 15 are not met.



39. D
The mean residual value for *all* regression analyses is 0. With a p-value of 0.0006, there is a significant relationship between X and Y, and the correlation is a high 0.911. A pattern in the residual plot, however, indicates that an even better fit exists.

40. D
Reversing the dependent and independent variables does not change the correlation. Since $R^2 = 0.795$ and the slope is positive,

$$r = \sqrt{0.795} \approx 0.892.$$